

ОТЗЫВ

официального оппонента на диссертационную работу
МОСИНА Сергея Владимировича «Методы и алгоритмы формирования многомерных данных с использованием промежуточных представлений», представленную на соискание ученой степени кандидата физико-математических наук по специальности 05.13.11 - «Математическое обеспечение вычислительных машин, комплексов и компьютерных сетей»

Актуальность темы диссертационной работы С.В. Мосина определяется тем, что в настоящее время происходит интенсивное накопление больших объемов данных в различных сферах производства. Во многих организациях стоит проблема обработки и анализа накопленной информации. Как в отечественной, так и в зарубежной литературе в качестве решения проблемы анализа накопленной информации традиционно используется технология оперативной аналитической обработки данных (*online analytical processing, OLAP*). Однако в рамках данной технологии обычно не принимается во внимание специфика прикладных задач, требующих частого обновления анализируемых данных в различных информационных контекстах. Кроме того, используемые подходы кэширования данных при построении гиперкубических представлений в основном опираются на результаты уже выполненных запросов, не используя возможности частичного или комбинированного использования текущего состояния данных. Все это приводит к существенным затратам времени и труда на построение новых многомерных представлений данных. Актуальной является задача разработки процессов динамического преобразования реляционной базы данных к многомерному представлению данных и последующей оптимизации этих процессов.

Целью диссертационного исследования С.В. Мосина является разработка методов и алгоритмов формирования многомерного представления данных из реляционного представления при наложении логических ограничений на размерности и при использовании сохраненных данных.

В диссертации Сергея Мосина эта цель достигнута, и получены следующие **основные результаты**:

1. разработан и аналитически исследован алгоритм направленного перебора для формирования контекстов;
2. разработан оптимизированный алгоритм формирования представления данных «Таблица Соединений»;
3. предложен и исследован оригинальный метод сравнения областей истинности логических ограничений при анализе сохраненных (кэшированных) данных;

4. разработаны алгоритмы повторного использования сохраненных данных и вычисления недостающих данных на основе сравнения областей истинности;
5. реализовано программное обеспечение, формирующее многомерное представление из исходного реляционного представления с использованием сохраненных данных;
6. проведены вычислительные эксперименты, подтверждающие эффективность предложенных подходов.

Достоверность полученных результатов подтверждена теоремами и утверждениями, снабженными строгими математическими доказательствами. Вычислительные эксперименты, проведенные в соответствии с общепринятыми стандартами, подтверждают теоретические построения.

Научной новизной проведенного исследования характеризуются разработанный автором оригинальный метод сравнения областей истинности запросов к реляционной базе данных, на основе которого сформулированы алгоритмы использования сохраненных данных, а также оптимизация имеющихся алгоритмов построения гиперкубического представления данных.

Теоретическая ценность работы состоит в том, что в ней приводится формальное описание методов сравнения логических ограничений, классификация различных комбинаций формул запросов, обоснование результатов, касающихся частичного включения естественных соединений реляционных отношений.

Практическая значимость заключается в том, что на базе предложенных методов и алгоритмов разработано программное обеспечение, осуществляющее формирование многомерного представления данных из исходного реляционного представления.

В качестве замечаний можно отметить следующее.

1. Сравнение логических формул, по видимому, предполагается осуществлять после преобразования SQL запросов к какому-то внутреннему виду, но нигде не поясняется, как именно происходит этот процесс и какова его трудоемкость.
2. К описанному в работе преобразованию вполне применимо распараллеливание по технологии GPU, однако автор нигде не упоминает о возможном использовании данной технологии для ускорения формирования гиперкубов.
3. При описании экспериментов в третьей главе следовало бы более четко указать условия экспериментов. В частности, в Oracle и в Microsoft SQL Server имеются настройки кэширования, которые могут улучшить производительность систем.

Указанные замечания не снижают общей значимости исследования, выполненного С.В. Мосиным.

Результаты диссертационной работы были опубликованы в рецензируемых научных изданиях, рекомендованных ВАК, и прошли апробацию на международных научных конференциях. Количество публикаций в рецензируемых научных изданиях соответствует требованиям Положения о порядке присуждения ученых степеней. Текст автореферата соответствует содержанию диссертации.

Диссертационное исследование С.В. Мосина является завершенной научно-квалификационной работой, в которой предложено решение задачи динамического построения многомерных баз данных из исходных реляционных баз данных с использованием сохраненных результатов запросов. Считаю, что диссертация Мосина С.В. соответствует критериям, установленным Положением о порядке присуждения ученых степеней, включая пункт 9 Положения, и является самостоятельным и завершенным научным исследованием, а Мосин С.В. заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Главный научный сотрудник Федерального государственного бюджетного учреждения науки Институт системного программирования им. В.П. Иванникова Российской академии наук,

доктор тех. наук, профессор

С.Д. Кузнецов

15 сентября 2017 г.

Электронная почта: kuzloc@ispras.ru

Тел.: +7 (916) 119-85-01

Адрес организации: Россия, 109004, г. Москва, ул. Александра Солженицына, 25

Подпись С.Д. Кузнецова заверяю:

ученый секретарь ИСП РАН, к.ф.-м.н.

Н.В. Пакулин

